SUPPLEMENTAL MATERIAL

PARTICIPATION IN OTHER SPORTS

Weightlifters train concurrently in other sports, of which CrossFit and endurance physical activities are the most common.

Reference:
Huebner M, Faber F, Currie K, *et al.* How Do Master Weightlifters Train? A Transnational Study of Weightlifting Training Practices and Concurrent Training. *International Journal of Environmental Research and Public Health* 2022;**19**:2708. doi:10.3390/ijerph19052708

Table S1: Concurrent training and prior sport participation

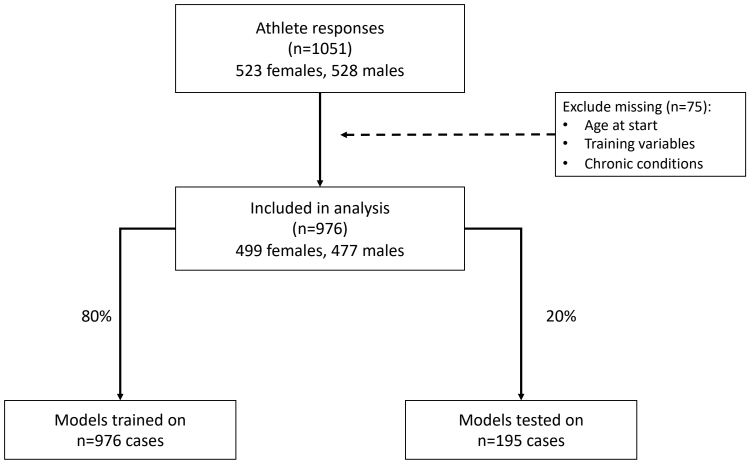
	Concurrent training		Prior sport participation	
	Females	**Males**	**Females**	**Males**
CrossFit	0.45 $^{225}/_{499}$	0.30 $^{141}/_{477}$	0.68 $^{341}/_{499}$	0.42 $^{202}/_{477}$
Ball sports	0.05 $^{27}/_{499}$	0.14 $^{65}/_{477}$	0.30 $^{149}/_{499}$	0.51 $^{243}/_{477}$
Martial arts	0.02 $^{10}/_{499}$	0.04 $^{19}/_{477}$	0.11 $^{53}/_{499}$	0.18 $^{85}/_{477}$
Body building	0.11 $^{54}/_{499}$	0.16 $^{76}/_{477}$	0.37 $^{186}/_{499}$	0.40 $^{191}/_{477}$
Powerlifting	0.06 $^{28}/_{499}$	0.07 $^{33}/_{477}$	0.13 $^{65}/_{499}$	0.21 $^{98}/_{477}$
Fitness	0.20 $^{101}/_{499}$	0.21 $^{101}/_{477}$	0.39 $^{194}/_{499}$	0.28 $^{133}/_{477}$
Endurance	0.24 $^{119}/_{499}$	0.28 $^{132}/_{477}$	0.45 $^{223}/_{499}$	0.33 $^{158}/_{477}$
Track and field	0.01 $^{7}/_{499}$	0.04 $^{21}/_{477}$	0.16 $^{80}/_{499}$	0.25 $^{118}/_{477}$
Yoga/Pilates	0.19 $^{94}/_{499}$	0.05 $^{24}/_{477}$	0.24 $^{118}/_{499}$	0.05 $^{25}/_{477}$
Gymnastics	0	0	0.07 $^{34}/_{499}$	0.02 $^{9}/_{477}$

METHODS FOR MACHINE LEARNING

The aim of is to predict a binary outcome, namely injury, as a function of covariates. For machine learning approaches the dataset is divided into a training set and then performance is evaluated in a validation or test set (ref Liu, Jama 2019). Machine learning models have prespecified settings, called hyperparameters. These hyperparameters regulate the trade-off between over-fitting and under-fitting a model. The optimal values of hyperparameters cannot be found by fitting the models with data. Tuning is the process to identify the value of the hyperparameter through searching among a series of values. We used a 10-fold repeated cross-validation to tune the parameter estimates for each ML algorithm.

A flow chart summarizing the number of athletes and exclusions due to missing training variables or chronic conditions is shown in Figure S1.

Figure S1. Study flow diagram for training and test data.



There were 41 variables divided into the following groups of predictors:

- demographic variables (p=5): sex, age, education level, age at start of weightlifting, and years of experience
- chronic conditions (p=5): high blood pressure, cardiovascular disease, cancer, diabetes, and arthritis/osteoarthritis
- training frequency (p=3): number of days per week, length of training session, hours per week (estimated)
- time for core weightlifting training (p=2): classic lifts (snatch, clean and jerk), strength exercises (squats, presses)
- training extensions (p=4): time for warm-up, cool-down, or supplementary exercises, attitude towards nutrition to support training (recovery, muscle gain)
- training program (p=3): following own or a coach's program (in-person or remote)
- concurrent training (p=9): CrossFit, ball sports, martial arts, body building, powerlifting, endurance training, general fitness, track and field, mobility (e.g. yoga/Pilates)
- prior sport participation (p=10): all concurrent sports and gymnastics

For each test set the models produce a probability of injury based on the covariates. Finally, the predictions were ensembled across the algorithms to combine information from each ML algorithm and possibly achieve better prediction performance for each injury location. This was repeated for all variables and leaving out the specified groups of variables. Performance was measured as accuracy which is the proportion of correctively predicted results out of the entire sample. We used the algorithms to classify the injury cases and ensemble the predictions for injuries at these locations. The performance of random forest and logistic regression models were used as references, using all variables and using a selection of variables.

The R packed caret was used to implement the algorithms, tuning the hyperparameters and ensemble the predictions.

Reference:
Kuhn M. Caret package. *Journal of Statistical Software* 2008; **28**(5)

OVERVIEW of machine learning algorithms used for the ensemble prediction models

1. Support vector machine (SVM)
   were developed by Cortes and Vapnik (1995) for classification and regression problems. For classification tasks in SVM, all data points are mapped in an n-dimensional space which is formed by the n features or covariates. Then the algorithm searches for an optimal decision hyperplane using the covariates to separate points to different classes. SVM uses a kernel function to project the input data space in certain form. There are several choices of kernel functions, we tried the most popular kernels including linear, radial, and polynomial kernels. The selection is decided by the accuracy of the holdout set. The c parameter controls the trade-off between the sensitivity of the boundary among groups and misclassification rate. When c is small, the penalty for misclassification is low the boundary is less sensitive to noise, but the misclassification rate will be higher. When c is large the effects are the opposite. We identified the best value of c by trying a range of values ("tuning of the c parameter").

   Reference:
   Cortes C, Vapnik V. Support-vector networks. *Machine Learning*. 1995; **20**: 273–297.

2. Stochastic Gradient Boosting is one of the ensemble methods, which ensemble predictions of a series of weak learning models, usually decision tree models. These models perform only a little better than random guessing, hence are called weak learning models. Each weak leaning model improves upon the previous model. Stochastic gradient descent samples a subset of data to grow the subsequent tree. This method helps the algorithm to avoid local minima and to approach the global minimum. We tuned several Tuning parameters are the number of trees, depth of each tree (how many branches to ensemble), learning rate (how quickly the algorithm proceeds down the gradient descent), and the minimum number of observations allowed in the trees' terminal nodes.

   Reference:
   Friedman JH. Stochastic Gradient Boosting, *Computational Statistics and Data Analysis* 2002; **38**(4):367-378.

3. Regularized logistic regression uses regularization to penalize model complexity. Regularization constrains the sizes of the coefficients. There are two main types of regularizations: Ridge regression and least absolute shrinkage and selection operator (LASSO). Both methods use a tuning parameter lambda to decide the importance of the penalty. Higher penalty reduces the magnitude of the coefficients. We used a hybrid model of Ridge and LASSO called Elastic Net, which linearly combines the two. The best combination is identified through cross-validation.

   Reference:

Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*,2010; **33** (1): 1–22. https: //doi.org/10.18637/jss.v033.i01.

Tibshirani R, Bien J, Friedman, J *et al*. 2012. Strong Rules for Discarding Predictors in Lasso-Type Problems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 2012;* **74** (2): 245–66. https://doi.org/10.1111/j.1467-9868.2011.01004.x.

4. Random forests is another type of ensemble method. It creates multiple classification trees by drawing bootstrapped samples of the data and randomly select subsets of the predictors. Then, it ensembles the predictions of the trees. Random forests can capture the nonlinear associations between predictors and outcome variables. We tuned the Tuning parameters are the number of randomly selected predictors, splitting rules (Gini index and `extratrees`), and minimal node size.

   Reference:
   Breiman, L. Random forests. *Machine Learning* 2001; **45**:5-32. 10.1023/A:1010933404324.
   Wright MN, Ziegler,A. ranger: A fast implementation of random forests for high dimensional data in C++ and R. *J Stat Softw* 2017; **77**:1-17. 10.18637/jss.v077.i01.

5. Single-hidden-layer neural network models can be viewed as nonlinear regression models. These models extract features (or units) from data and then come up with predictions. The complexity of the models is partially dependent on the number of layers. Each layer can be considered as one time of feature extraction. The number of features is determined by the number of units. The type of neural networks that are considered in this paper is the single-hidden-layer neural network. This function allows the users to adjust the number of units and the regularization parameter to avoid overfitting. We tuned two hyperparameters: Tuning parameters are number of hidden units and decay, which are used for avoiding over-fitting.

   Reference:
   Venables WN, Ripley BD(2002) *Modern Applied Statistics with S*. Fourth edition. Springer.

6. Naïve Bayes is a classification method employing Bayes' rule to estimate the conditional distribution of the input variables given the value of the outcome variable. Naïve Bayes assumes conditional independence of the input variables given the outcome class. The algorithm allows the users to add Laplace smoother to avoid zero posterior probability. Users can also adjust the how flexible density estimate is. The tuning parameters included Laplace Correction, and Bandwidth Adjustment. The Laplace correction is a solution of zero probability problem for test data. The bandwidth controls the spread in kernel density estimates function.

   Reference:

Zhang Z. Naïve Bayes classification in R. *Ann Transl Med* 2016; **4**(12):241. doi: 10.21037/atm.2016.03.38

Ensembling combines information from multiple machine learning models to improve predictive accuracy. We used a generalized linear model to create a simple linear blend of models. It calculates weighted averages of variable importance for each model.

TABLE S2. Number and proportion of injuries at different locations in training and test sets

Injury location	**Training set, n (%)**	**Test set, n (%)**
Shoulders	277 (35.4)	66 (34.0)
Knees	204 (26.1)	54 (27.8)
Back	184 (23.5)	45 (23.2)
Wrists	172 (22.0)	39 (20.1)
Hips	103 (13.2)	20 (10.3)

RESULTS FOR MACHINE LEARNING MODELS

The accuracy of the ensemble model was 0.773 (back), 0.727 (knees), 0.644 (shoulders), 0.774 (wrists), 0.876 (hips) (Table S3). Leaving out groups of variables (demographics, training, concurrent, or prior sports) did not appreciably lower the accuracy. This could be explained that at least one variable in each group was ranked among those with high importance metric. Model predictions of shoulder injuries were less accurate compared to knee, back, or wrist injuries.

Random forest models alone had similar or better accuracy when using the expert selected subgroup of variables, with 0.748 (back), 0.713 (knees), 0.683 (shoulders), 0.744 (wrists), 0.901 (hips) (Table S3).

Table S3. Accuracy of performance for ensemble predictions and groups of variables

	Back injuries	Knee injuries	Shoulder injuries	Wrist injuries	Hip injuries
All variables	0.773	0.727	0.644	0.774	0.876
Without demographics	0.749	0.749	0.696	0.774	0.899
Without chronic conditions	0.737	0.716	0.655	0.841	0.866
Without training frequency	0.727	0.687	0.677	0.789	0.864
Without time for core weightlifting training	0.769	0.733	0.651	0.776	0.877
Without training extension	0.785	0.712	0.659	0.762	0.883
Without training program	0.764	0.733	0.662	0.776	0.872
Without current sports	0.763	0.722	0.639	0.759	0.876
Without prior sports	0.773	0.711	0.660	0.780	0.876
Reference: Random forest, all variables	0.768	0.742	0.619	0.790	0.876

Reference: logistic regression, all variables	0.737	0.711	0.613	0.759	0.876
Ref Random forest, selected variables	0.748	0.713	0.683	0.744	0.901
Ref logistic regression, selected variables	0.564	0.609	0.620	0.647	0.643

Due to the nonlinear prediction functions, it is not straightforward to interpret the results of ML algorithms. However, it is possible to obtain a variable importance metric (VIM) to ascertain the relative contribution of a particular covariate to the accuracy of the prediction. The exact method of calculating the variable importance often depends on the algorithm in use. For example, if the algorithm is linear regression, the importance is the absolute value of the t-statistic for each variable in the model. The R package, caretEnsemble, provides a weighted estimate of the importance of variables of a series of algorithms used. Variable importance does not indicate whether a variable is a risk factor for injuries or might contribute to prevention, it only indicates the relative contribution to prediction power of a variable.

Reference:
Strobl C, Boulesteix AL, Kneib T, *et al.* Conditional variable importance for random forests. *BMC Bioinform* 2008; **9**: 307.

The variables importance measures for the injury locations are shown in Table S4 and Figures S3 A-E.

Table S4. Variable importance measures for the ensemble method

Variable	Back injuries	Knees injury	Shoulder injuries	Wrist injuries	Hip injuries
Sex	0.0201	2.1675	8.9561	0.4699	0.0000
Age	1.5027	0.7611	6.4249	7.9517	5.4208
Education level	2.5486	0.7719	2.0882	2.2403	5.0546
Age at start of weightlifting	6.9056	7.9342	3.9839	4.5010	1.6997
Years of experience	7.5524	6.1908	6.0684	0.4628	0.6879
Number of training days	0.0150	1.0462	2.1764	4.8406	7.4276
Length of training session	0.6902	4.3760	1.3179	5.6153	2.2368
Hours per week (estimated)	0.5224	2.3640	2.7262	7.8245	7.0177
Time warm-up	1.1213	0.8427	3.1203	0.4793	4.9092
Time for classic lifts	1.5705	0.9438	1.8379	3.9940	0.5399
Time for strength training	2.3268	2.8536	2.4605	2.7149	2.4790
Training supplementary	3.4117	0.5477	1.8633	2.1798	5.3947
Time cool down	0.5126	0.4109	5.4972	0.1788	2.0645
Program coach	5.4603	5.2367	0.9530	2.4628	3.8571
Program remote coach	0.9236	1.0850	0.4139	5.4796	2.0420
Program own	6.8504	5.3986	2.3675	1.8715	3.6771
Nutrition	1.0233	2.4454	3.6426	0.5174	0.4557

Prior powerlifting	2.9224	1.2432	0.3769	0.0736	2.9948
Prior body building	4.7814	2.8995	1.1029	1.9375	1.9689
Prior CrossFit	4.9925	1.7108	1.5094	4.5809	4.9684
Prior endurance training	0.7614	1.0079	1.0033	0.3228	0.0000
Prior track& field	2.2164	1.6686	0.5051	1.9601	1.7163
Prior ball sports	2.7812	3.6487	3.3653	5.2771	3.2946
Prior fitness training	0.0093	0.6097	1.0763	0.9928	3.5728
Prior martial arts	2.3335	4.3317	1.1520	2.5985	1.9946
Prior yoga/Pilates	1.4434	2.5184	0.8633	0.2614	3.3285
Prior gymnastics	1.5823	1.0371	0.3183	0.2130	0.5750
Current powerlifting	1.0384	0.6188	0.5111	0.8423	1.9794
Current body building	1.4878	0.8020	5.6467	1.7056	3.5829
Current CrossFit	4.7338	4.5291	2.6723	0.7968	0.6809
Current endurance training	0.9502	1.9974	0.5935	2.7244	0.4411
Current track & field	0.0281	0.3657	0.3752	1.5516	0.2093
Current ball sports	0.0304	0.5015	0.4743	3.1253	0.8678
Current fitness training	1.5137	1.2812	4.6571	0.4202	0.8590
Current martial arts	1.8218	1.6803	0.0142	0.8301	0.3289
Current Yoga/Pilates	1.9394	1.4955	0.3201	0.1092	0.7106
Chronic inflammation/osteoarthritis	0.0156	2.1485	8.5201	0.1611	0.0004
High blood pressure	0.0046	0.0370	1.1085	0.1175	0.0004
Cardiovascular diseases	0.0000	0.6192	0.0700	0.0421	0.0001
Cancer	0.0069	0.4201	0.0965	0.0534	0.0000
Diabetes	0.0111	0.8562	0.1002	0.0377	0.0001

Figure S3A. Variable importance measure for shoulder injuries

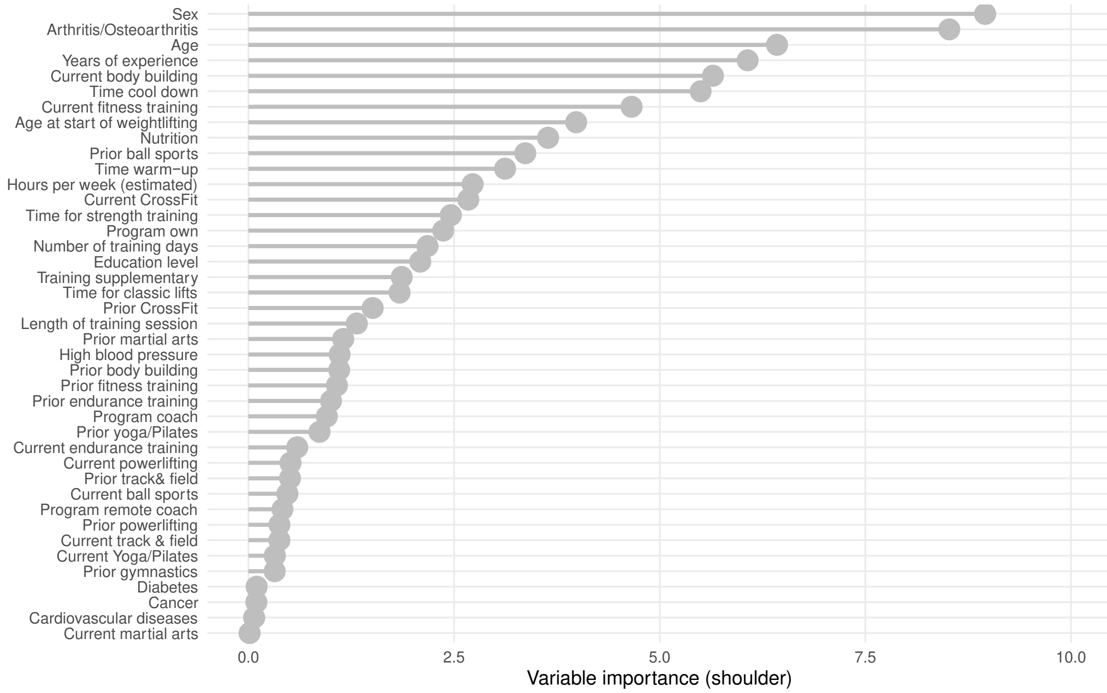


Figure S3B. Variable importance measure for back injuries

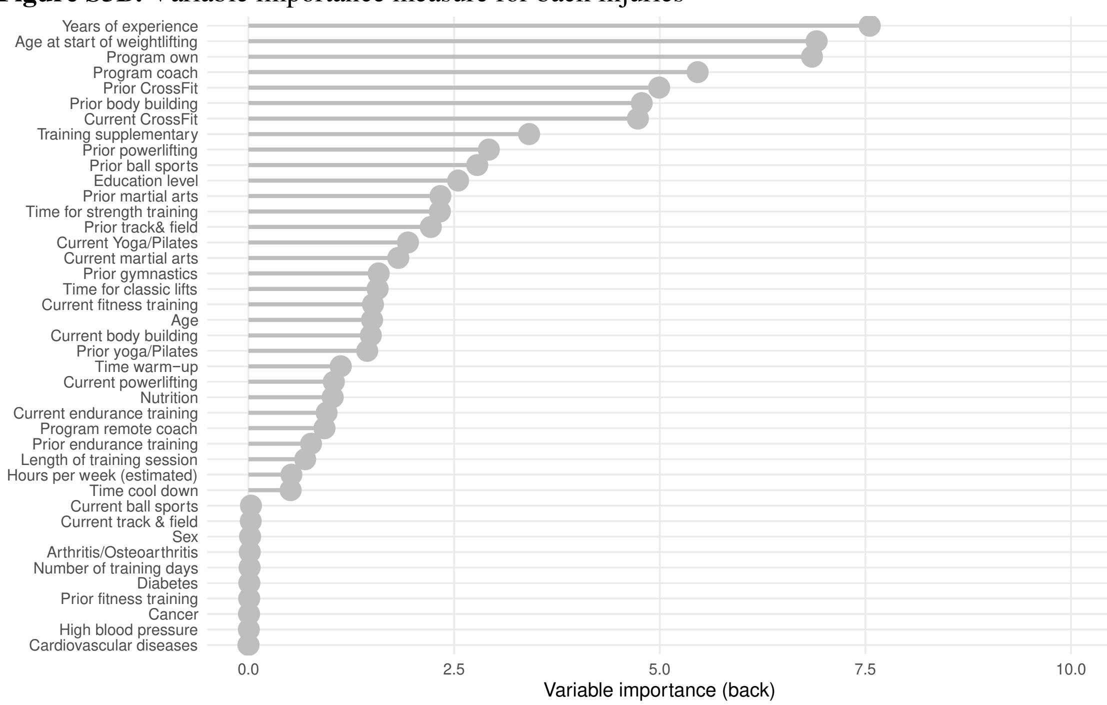


Figure S3C. Variable importance measure for hip injuries

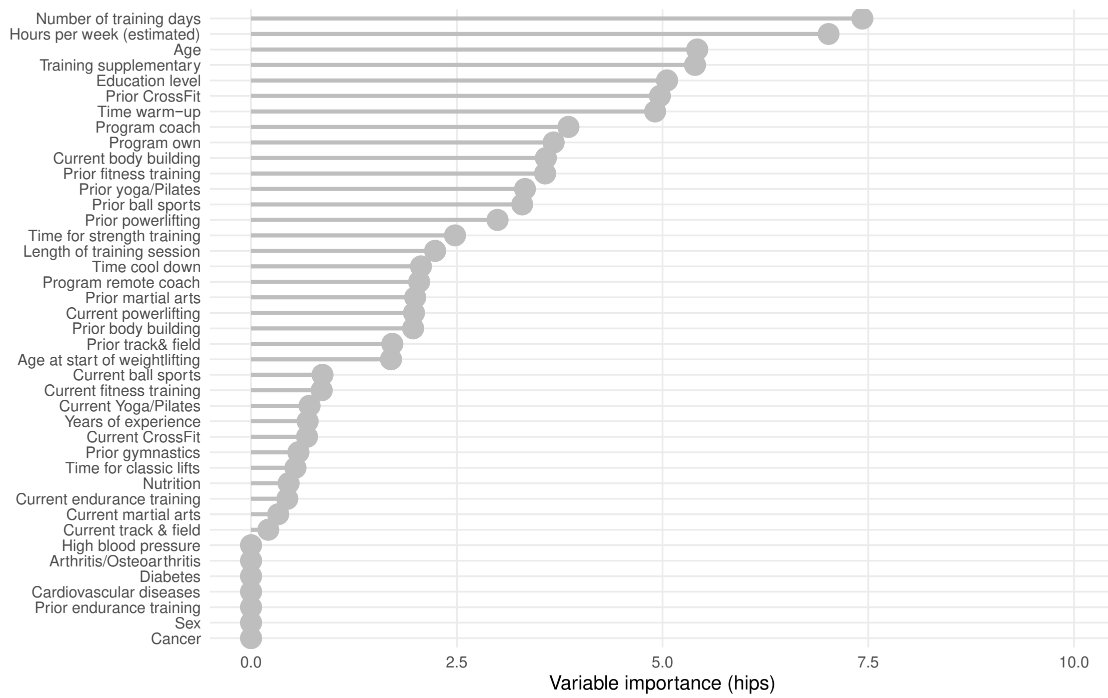


Figure S3D. Variable importance measure for knee injuries

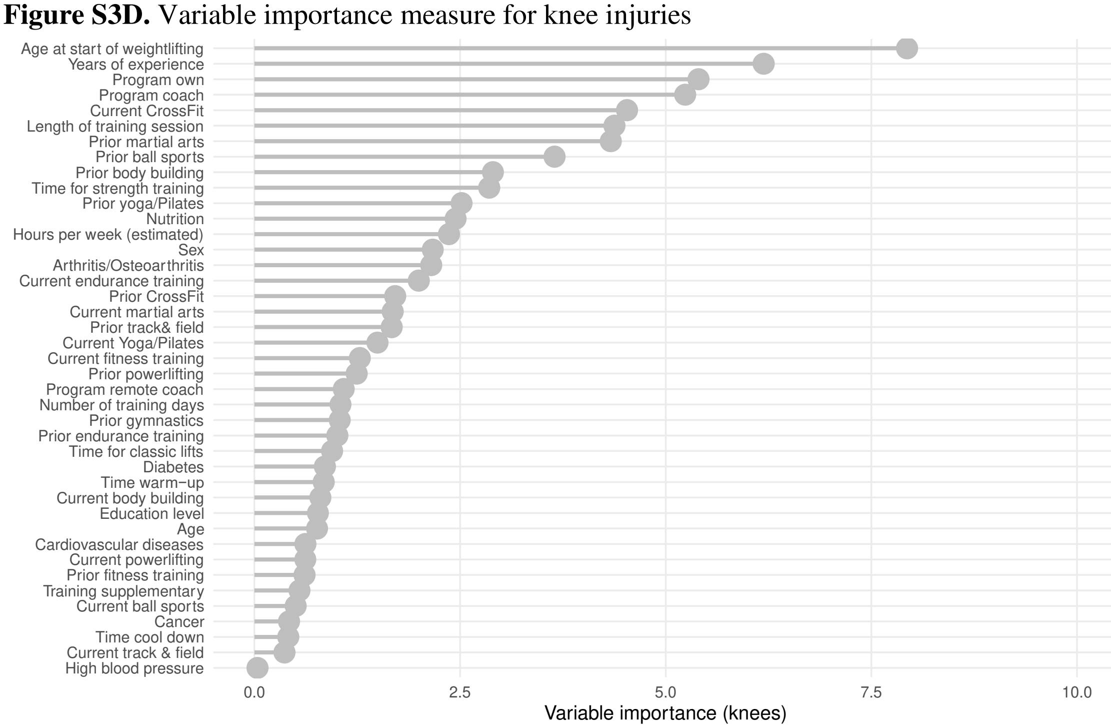
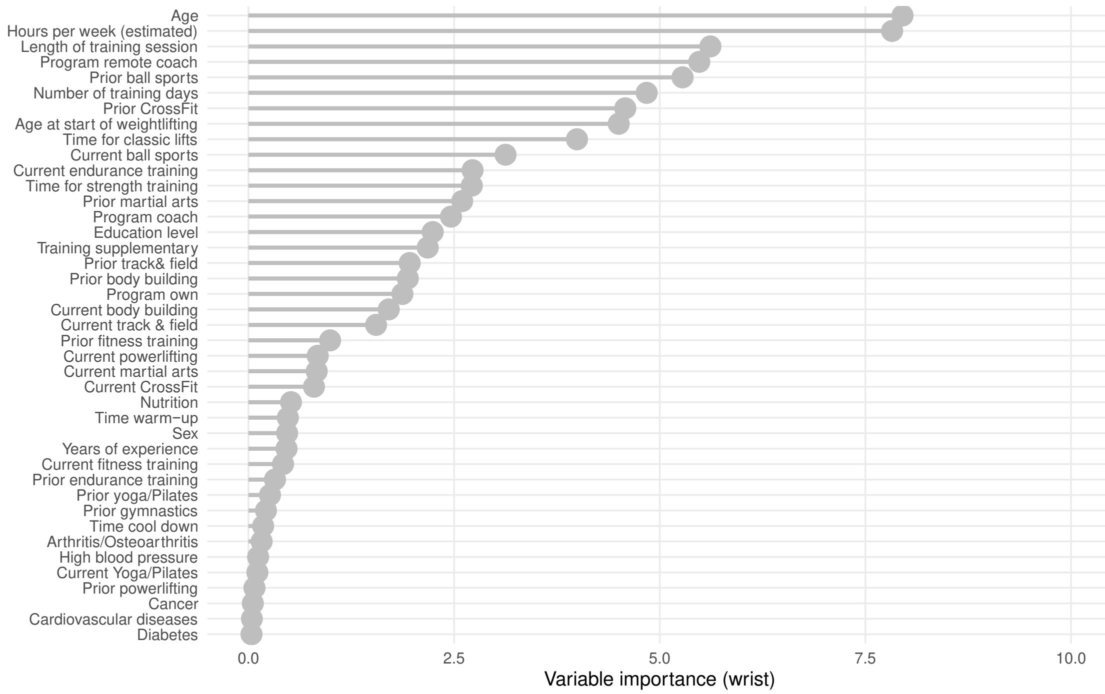


Figure S3E. Variable importance measure for wrist injuries



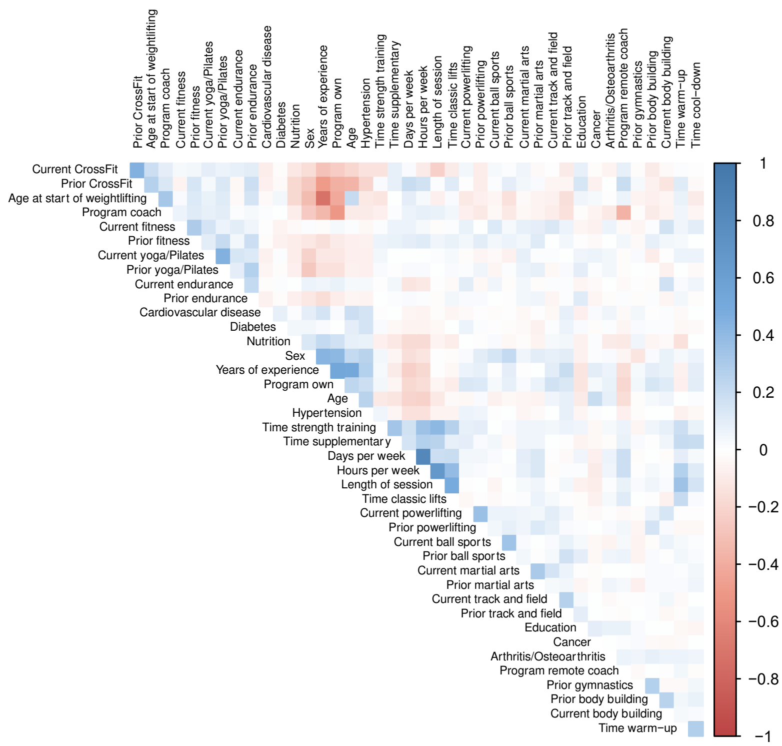
EXPERT MODEL

The selected variables comprised sex, age, nutrition, chronic inflammation, and training variables such as days per week, time for warm-up, time for cool down, and time for supplementary exercises. Prior sports (body building, power lifting, ball sports, gymnastics) were combined (1= prior participation in any vs 0 = no prior participation). Concurrent mobility such as yoga or Pilates and concurrent participation in Crossfit were also selected. Due to the collinearity of years of experience and age at start of weightlifting, the latter was included in the model since it was mentioned more often. Interactions between variables were mentioned by the experts, but did not improve the model fit as measured by the concordance statistic.

CORRELATIONS

Correlation coefficients of absolute values 0.5 or higher are in derived variables such as number of training days and with hours per week or age at start of weightlifting and years of experience.

Figure S2. Correlation plot of predictor variables.



SURVEY QUESTIONS

What is your current age?

Have you participated in sports or physical activities before you started weightlifting?

- ○ Yes
- ○ No

→ If yes, which activities/sports have you participated in? __________

- Bodybuilding
- Powerlifting
- CrossFit
- Fitness
- Endurance sport (e.g. running, swimming, cycling, skiing, hiking)
- Track and Field
- Martial arts
- Ball sports
- Pilates/Yoga
- Other, please specify: ____________________

In a typical week have you also participated in the following physical activities/sport in addition to weightlifting before the pandemic? (click all that apply)

- Bodybuilding
- Powerlifting
- CrossFit
- Fitness
- Endurance sport (e.g. running, swimming, cycling, skiing, hiking)
- Track and Field
- Martial arts
- Ball sports
- Pilates/Yoga
- Other, please specify: ____________________

What program(s) were you following in your weightlifting training before the pandemic? (click all that apply)

- The program assigned by my in-person coach.
- The program assigned by my remote coach.
- My own program
- A program from a website/book/subscription
- Other (please specify) ____________

In a typical week on how many days have you trained in weightlifting prior to the pandemic?

- 1 day
- 2 days
- 3 days
- 4 days
- 5 days
- 6 days
- 7 days

How long was a typical weightlifting training session for you including warm-up and cool-down prior to the pandemic?

- <1 hour
- 1-<1.5 hours
- 1.5-<2 hours
- ≥2 hours

How long was your typical warm-up before the pandemic?

- 0-<15 minutes
- 15-<30 minutes
- ≥30 minutes

On average how much time in your typical training session before the pandemic was devoted to the competition lifts (snatch, clean & jerk) and partial competitions lifts (such as hang snatch or clean from blocks)?

- 0-<15 minutes
- 15-<30 minutes
- 30-<45 minutes
- 45-<60 minutes
- ≥60 minutes

On average how much time in your typical training session was devoted to strength exercises (squats, presses) before the pandemic?

- 0-<15 minutes
- 15-<30 minutes
- 30-<45 minutes
- 45-<60 minutes
- ≥60 minutes

On average how much time in your training session was devoted to additional exercises prior to the pandemic? (pull-ups, core, GHD, machines, etc)

- 0-<15 minutes
- 15-<30 minutes
- 30-<45 minutes
- 45-<60 minutes
- ≥60 minutes

How long was your typical cool-down (stretching, cycling, rowing,,...)

- 0-<15 minutes
- 15-<30 minutes
- ≥30 minutes

Have you ever had training restrictions due to acute injuries **during weightlifting** (click all that apply)?

- Shoulder joints
- Elbow
- Wrist
- Hips
- Knees
- Ankles
- Spine/back
- Bone, muscle, or tendon injuries (Please specify) _____
- no acute injuries

Have you ever experienced chronic inflammation or wear and tear?

- Yes (Please specify) _________
- No